SCALABLE CLUSTERING FOR VISION USING GPUs

K Wasif Mohiuddin

200807007

Abstract

Clustering algorithms have wide applications in Computer Vision, Data mining, Data Visualization, etc. Clustering is an important step for indexing and searching of documents, images, video, etc. Clustering large numbers of high-dimensional vectors is very computation intensive. CPUs are unable to handle such load and consume sometimes days and even weeks to cluster large data. GPUs are being used for general purpose computing of wide range of problems which require high computational power. Today's GPUs deliver as high as 1.5 TFLOPs. The GPU has evolved over the time not only by increasing number of cores but also major architectural changes like faster access of data, more shared memory, threads per block, etc. Such changes have enabled the GPU programmers to exploit its architectural features to the fullest and achieve high performance. In this thesis, we focus on K-Means algorithm which is a widely used unsupervised clustering algorithm. We develop a GPU based K-Means implementation for large datasets; also we have used this implementation to develop a video organizing application.

We present the design and implementation of the K-Means clustering algorithm on the modern GPU. All steps are performed entirely on the GPU efficiently in our approach. We also present a load balanced Multi-node, Multi-GPU implementation which can handle up to 6 million, 128-dimensional vectors. We use efficient memory layout for all steps to get high performance. The GPU accelerators are now present on high-end workstations and low-end laptops. Scalability in the number and dimensionality of the vectors, the number of clusters, as well as in the number of cores available for processing are important for usability to different users. Our implementation scales linearly or near-linearly with different problem parameters. We achieve up to 2 times increase in speed compared to the best GPU implementation for K-Means on a single GPU. We obtain a speed up of upto 170 on a single Nvidia Fermi GPU compared to a standard sequential implementation. We are able to execute single iteration of K-Means in 136 seconds on off-the-shelf GPUs to cluster 6 million vectors of 128 dimensions into 4K clusters and in 2.5 seconds to cluster 125K vectors of 128 dimensions into 2K clusters.

Video data is increasing rapidly along with the capacity of storage devices owned by a lay user. Users have moderate to large personal collections of videos and would like to keep them in an organized manner based on its content. Video organizing tools for personal users are way behind even the primitive image organizing tools. We present a mechanism in this thesis to help ordinary users organize their personal collection of videos based on categories they choose. We cluster the PHOG features extracted from selected key frames using K-Means to form a representation for each user-selected category during the learning phase. During the organization phase, labels from a K-NN classifier on these

cluster centers for each key frame are aggregated to give a label to the video while categorizing. Video processing is computationally intensive. To perform the computationally intensive steps involved, we exploit the CPU as well as the GPU that is common even on personal systems. Effective use of the parallel hardware on the system is the only way to make the tool scale reasonably to large collections that will be available soon. Our tool is able to organize a set of 100 sport videos of total duration of 1375 minutes in about 9.5 minutes. The process of learning the categories from 12 annotated videos of duration 165 minutes took 75 seconds on a GTX 580 card. These were on a standard desktop with an off-the-shelf GPU. The labeling accuracy is about 96% on all videos.

The ideas, approaches proposed in this thesis have been implemented and validated with experimental results. For large data-sets we developed a scalable, efficient K-Means clustering on GPU along with a Multi GPU framework and used it to develop a video organizer application providing high accuracy.